MANUAL

FOR THE

Purdue Rating Scale

FOR

Instructors

BY

G. C. BRANDENBURG

AND

H. H. REMMERS

1928

DUKE UNIVERSITY LIBRARY
DURHAM, N. C.

Rec'd March 26, 1929
Summer School
Fund

MANUAL

for the

Purdue Rating Scale

for

Instructors

By
G. C. Brandenburg, Ph. D.
Professor of Education and Psychology
Purdue University
and
H. H. Remmers, Ph. D.
Assistant Professor of Education and Psychology
Purdue University

P20573

Published by
Lafayette Printing Company
Lafayette, Indiana

Digitized by the Internet Archive
in 2020 with funding from
Duke University Libraries

https://archive.org/details/manualforpurduer00bran

371.1
T 817

PART I

GENERAL CONSIDERATIONS

P 20573

THE PURDUE RATING SCALE FOR INSTRUCTORS

FOREWORD

Not so very long ago it was believed by many educators that if a teacher knew his subject he could teach successfully. They contended that attention to the matter of the method of presentation constituted misdirected effort and was a sheer waste of time. And such a view is still by no means uncommon. Fortunately it is not so general as formerly. It is entirely possible that in some quarters altogether too much emphasis has been placed on other aspects of teaching than a thorough knowledge of the subject. Indeed it is quite certain that this not only has been true but still is true in some of our teacher training institutions. This is unfortunate. It must not be used, however, as an argument against the devotion of a reasonable part of the teacher's attention to those aspects of his work which are usually subsumed under the title Technique of Teaching. It is axiomatic that thorough knowledge of subject-matter is the first and most fundamental prerequisite of successful teaching. Very little observation is necessary, however, to convince one that such thorough knowledge of subject-matter is no guarantee of ability to teach.

The authors of the Purdue Rating Scale for Instructors believe that some such device as is described in this little booklet may become a valuable instrument for self improvement in classroom procedure in the hands of an intelligent teacher. They are firmly convinced that only as the teacher be-

comes independently aware of opportunities for self-improvement will he make the most of them. We are so constituted that we resent outside supervision, however constructive and competent the criticisms of the supervisor may be. We prefer to be our own supervisors, particularly with reference to such subjective qualities as the personality traits involved in this rating scheme. And this is, quite likely, as it should be. There is a strong case to be made for self-reliance, self-respect, and initiative on the part of the classroom teacher. While it is not to be understood that the authors are opposed to professionally competent supervision, as education is now organized it is probably just about as likely that a competent teacher will find himself under the supervision of a less competent one as is the reverse. Like any similar device The Purdue Rating Scale must be used with discretion and intelligent interpretation.

This booklet, which is largely a reprint of two articles in Educational Administration and Supervision, September and November, 1927, describes the construction and the methods of administering and interpreting the Purdue Rating Scale. If it results in ever so little increase in the tendency of the teacher to examine in a critical and constructive manner his classroom procedure it will have accomplished its purpose.

THE AUTHORS.

RATING SCALES FOR INSTRUCTORS

What do your students really think of you as an instructor? Do you know or do you care to know? Perhaps you are happier in your work if you do not know. Or possibly you are one of those "students be damned" instructors to whom it would make no difference. In many cases it will be granted that the student's interest in, and attitude toward a given course is conditioned largely by his personal reactions to the instructor. Whether the student's judgment is correct or not is another matter. The point is that his judgment of his instructor is a factor in itself in the total learning situation. Moreover, it is a most vital factor. Probably aside from the learner's general intelligence it is the most important element in the situation. For some time we have been experimenting with a device through which student judgments of instructors might be obtained in a convenient, systematic, and standardized form. Instructors who have used the rating scheme described in this booklet have encouraged us to believe that it may become a most useful instrument in the hands of the teacher who is really interested in self-improvement.

Basic Principles in Selecting Traits

In the selection of traits to be used in such a rating scheme as is here described three important principles should be observed: (1) The number should be relatively small so that the student shall not be lost in the maze of different qualities. The well known "halo" effect which is partly due to too long a list of traits we have found to be almost

wholly absent in our rating scale. The actual coefficient of correlation between any two given traits is found to be in most cases unusually low in comparison with published data on other rating scales. The list of traits must also be short enough that the student shall not be bored to the point of carelessness by the process. (2) The traits must be such as are generally agreed upon by competent critics as most important or at least among the most important ones. (3) The traits must be of such nature that they are fairly susceptible to student observation and judgment.

Number of Traits Used

It has been found through considerable experimentation that it is not desirable to use in such rating schemes more than about ten separate items or traits. The ten traits which appear in our rating scale have been selected from various sources. They are all found under one name or another in the list of desirable qualities for instructors found in the report of the "Better Yet" Faculty-student Committee in the Quality of Instruction in Elementary Courses, University of Chicago, published in the Bulletin of the American Association of University Professors, April, 1926.

Are Student Judgments Reliable and Valid?

The best measure of the degree to which these traits admit of reliable observation and judgment by the student is to be found in the reliability coefficients, and these have been found to be quite satisfactory. In a later section detailed evidence on this point will be presented.

We do not know how much validity the composite judgment of an instructor's students possesses as a true measure of the instructor's ability. We should certainly hesitate to accept the undergraduate's judgment of the value of an instructor as a whole to the university and to society in general. For undoubtedly there are many instructors who would be judged failures in the classroom by any competent judge and yet whose value to the university, or to the state, in one way or another is beyond question. As previously stated, the teacher who uses the scale must constantly bear in mind that it is primarily a device for ascertaining student judgment concerning the traits in question. Regardless of what the teacher believes or knows about himself with respect to these traits, the student attitude exists, and exists as a functioning factor in the classroom procedure.

While we are chiefly concerned here with the description of a technique for ascertaining student judgment because of its intrinsic value as a factor in the learning process, still we shall show in later sections of the booklet that such judgment does possess considerable validity.

Ratings Must Be Anonymous

In order to secure judgments which shall be approximately true measures of the student's attitudes it is necessary to take such precautions as will guarantee the student that his rating will not react in any way on his standing with the instructor, for intelligent students are always on the lookout to discover how and where to invest their judgments so as to insure the maximum scholastic returns.

The "fifteen-minute-a-day-preparation" student who is always expressing his great interest in your course exemplifies the principle. The technique herein described frees the student quite thoroughly, we believe, from any feeling that his rating of the instructor can possibly militate for or against him in the course.

How to Proceed

In the first place the whole rating process may be conducted by a member of the class with the instructor outside of the classroom. If the instructor administers the test personally he should simply hand out the rating sheets one to each student with the request that the students read the directions and do what they tell him to do. In the second place, no names, numbers, or other identification marks should be placed on the rating sheets; and finally, no writing whatever should appear by which the individual student might be identified. It seems quite essential to provide such drastic precautions as those just mentioned in order to secure frank, sincere, and unbiased judgments where the student-instructor relationship is involved.

Are the Results Beneficial?

The question of the effect on the instructor's classroom work is naturally the most important one in connection with such a procedure. It is reasonable to assume that an instructor who would voluntarily undertake to secure such ratings would profit by the results. Observations thus far, while limited, support this assumption. Our experience with the scale leads us to believe that an instructor may have certain defects in his classroom teach-

ing and yet work day in and day out in blissful ignorance of their existence.

Two Cases

For example, A Mr. X who is known to his colleagues for his pride in presentation of his subject in the classroom was quite surprised to find that his students rated him lower in this particular than in any other aspect of his teaching. Moreover he discovered by comparison of his own ratings with those of his departmental colleagues made by the same students in the same course that he was rated relatively low in this trait. He now has a lively interest in methods of teaching. Another instructor whose unique mannerisms are so marked as to be frefuently commented upon by his associates was amazed to discover through the ratings by his students that his classroom behavior was not of the most conventional type. Of course, it must not for a moment be inferred that the exhibition by an instructor of strong or striking individuality through certain unusual behavior patterns detracts in itself from the effectiveness of his classroom teaching. It is altogether likely that in certain cases, individual peculiarities in procedure or what might even be regarded as striking mannerisms tend to increase the effectiveness of the instructor's presentation through increased attention on the part of the class. The point is that by means of the rating process the instructor is better able to evaluate the effect of his mannerisms on the learning and teaching process.

Use of the Scale Should Be Voluntary

While heads of departments and other executives might make considerable use of students' ratings in forming judgments of instructors, still the wisdom of forcing any instructor to submit to such procedures is extremely doubtful. In fact such a compulsory procedure is almost certain to defeat its own ends and result in more harm than good. The entire results of the rating should remain in the hands of the instructor to use as he sees fit. If comparisons are made among the ratings of various instructors the matter should be entirely voluntary on the part of all instructors concerned.

Some Desirable Effects

We have found it entirely feasible to allow a student or a committee of students from the class in which the rating is done to work up the results. In some cases a report has been made to the class. It might be supposed that this would cause undesirable discussion of the instructor by his students and the creation of a hypercritical attitude. This can hardly be urged as an objection, however, since in any case students openly discuss their instructors. In fact, there is reason to believe that the attitude of a class is generally improved by such a frank and open willingness on the part of the instructor to allow himself to be critically and publicly evaluated by his students. Whatever method of scoring is adopted, the instructor will find it worth while to make a careful study of the results for his own use. It is readily conceivable that an instructor may by using such a rating scheme in a number of succeeding classes improve his classroom procedure

and this is after all the ultimate aim. Of course such improvement as might result would be based on the assumption that students' ratings in the main have some degree of validity. And this, we shall show, is not a dangerous assumption.

PART II

EXPERIMENTAL DATA

EXPERIMENTAL DATA ON THE PURDUE RATING SCALE FOR INSTRUCTORS

I—VALIDITY

With the use of any measuring instrument the question always arises, Does it measure what it purports to measure? With standardized tests or with the new type of "objective" classroom tests the asking of this question not infrequently leads to a serious indictment of the test on the score of validity, for it can often be shown that the test in question is no more a test of arithmetic, geography, history, psychology, or what not, than it is a test of reading speed or comprehension, speed of handwriting, or of ability to follow directions.

The Purdue Rating Scale for Instructors is not open to this general criticism. It purports to measure the students' judgments of the instructor on the 10 categories given on the Scale. To the extent that students agree among themselves on the amount of the trait which the instructor possesses, and, more important, the extent to which each student is self-consistent in his judgments, to that extent we are in a position to say that the scale is valid.[1] For this scale reliability is snyonymous with validity.

II—RELIABILITY

In order to test the reliability of the scale, the procedure outlined below was followed. Two forms

[1]The theoretical logic for this position may be stated in the words of T. L. Kelley: If competent judges appraise Individual A as heing as much better than Individual B as Individual B is better than Individual C, then it is so, as there is no higher authority to appeal to. "The Influence of Nurture upon Individual Differences," Macmillan Co., 1926, p. 9.

of the scale were prepared, one in which the traits were "staggered," i. e., in which the zero end of the scale and the 100 per cent end of the scale were placed according to chance either at the left or the right margin of the paper, and a second form in which the zero end of each trait was placed at the right. The students who participated in this part of the investigational procedure were asked to rate the instructor on one form of the scale, without their being informed of the existence of the second form. Three days later when they had presumably forgotten specific judgments on the traits, they were again asked to rate the same instructor on the second form. In order to preserve the anonymity of the ratings while yet making it possible to correlate one form with the other, they had been asked to appoint a class secretary to assign to each student a number. These numbers he announced to the students in the absence of the instructor. On the occasion of the second rating the instructor again left the room while the secretary made sure that each student had his proper number so that his two ratings could be identified. This made possible a correlation without the instructor's knowing who any individual rater was in terms of his number. In order to divert the students' attention from the real purpose of this part of the study —i. e., a retest—they had been told at the time of the first rating that the class secretary would correlate ratings with mental test scores, which were already available. The students were unanimous in their testimony that they were unable to recall how they had checked the various traits on the first form when they were filling out the second.

The above procedure was gone through to discover the effect "staggering" would have upon the judgments—whether, as has been claimed, the "staggered" form would differ significantly from the "straight" form, in that such staggering would force greater discrimination on the part of the rater by preventing a general mental set which would lead him to rate the instructor about the same for all the traits. Our investigation on this point gave wholly negative results. As much discrimination was achieved by one form as by the other. Nor is there the slightest evidence that a general mental set influenced the results on the "straight" form. Any such constant error would necessarily manifest itself by increasing the mean values of traits on which a given instructor was rated relatively low. Reference to Tables IV and V will show that no such consistent differences as this theory requires are present. The absolute values for the means of traits 1, 2, 4, 6, and 7 are higher for the "straight" form, while those for traits 3, 5, 8, 9, and 10 are lower. The statistically most significant differences, i. e., traits 3 and 5, are indeed contrary to the mental set theory in that the values for the "straight" form are actually less than for the "staggered" form. Indeed, our results lead us to suspect that the alleged effect of the "straight" arrangement of rating scales in general was arrived at on *a priori* grounds rather than on the basis of experimental data. It is possible that there exists here a real difference between numerical rating scales and graphic rating scales. The point is perhaps worth investigating, but we have not done so.

The Number of Judgments Required

One of the objections frequently raised with reference to the use of rating scales is that the number of judgments obtained is too small, and results therefore unreliable. One of the writers has shown in connection with another rating scale that the Spearman-Brown prophecy formula applies with reasonable accuracy to the problem.[1]

We have made no specific investigation of this problem in connection with the present study, but the logic of the situation is in favor of assuming that the formula applies even better to results on the Purdue Rating Scale for Instructors, for the students in this case are probably more competent after a semester's acquaintance to judge the instructor on the 10 traits of our Scale than is the case in the study referred to above. If the Spearman-Brown formula applies, then in order to obtain a reliability of + .95 for the sum of all the traits under the conditions obtaining for the data of Table 1, Instructor T, approximately 120 student judgments are required. It follows from this that the pooled judgments of all the students in the various classes of a given instructor will give an extremely significant and stable index of his relative merit as an instructor,[2] in so far as student judgments are an index of this sort.

---

[1]Remmers, H. H. Shock, N. W., and Kelly, E. L.: An Empirical Study of the Spearman-Brown Prophecy Formula. Journal Educational Psychology, Mar., 1927.

[2]For a "common sense" corroboration of this conclusion see an article by Jordan, C. B.: A Rating Scale for Measuring the Professor. The Druggist Circular, Jan., 1927, pp. 223-224.

How Dependable Is the Mean?

The PEm as listed in the various tables is an index of the expected variability of the arithmetic mean with comparable samplings. For example, it is an even wager that trait 1, Table IV, column 1, will with further sampling receive a mean value as high as 91.306 or as low as 89.494.

The Agreement of Student Ratings with Other Judgments

Table IV gives data to show the discriminatory powers of the scale for a trait rumored to be particularly marked in the case of Instructor D, namely, personal peculiarities. This man is judged by students and colleagues to be an excellent teacher who according to his own testimony uses bizarre peculiarities as attention-getting devices. The mean values for trait 8 show significant depressions relative to other traits.

Should Judgments of the Same Teacher Vary from Class to Class?

That class judgments may differ significantly for the same man is brought out by Table VII, in which the differences for the totals of Table VI are studied for their significance. Classes II and III are both more than 2σ higher than class I. Such differences will never occur by chance.

Just what the reasons are for such variations in any particular case would be difficult to say. A large number of possible factors might enter in. It is quite possible, for example, that the higher ratings are obtained from a class which is more mature—seniors as against freshmen. The reverse,

of course might be equally true, i. e., that mature students might also be more critical. Again, higher ratings might be obtained if the students during the previous period had been under the instruction of a teacher whom they would rate very low. The contrast in such a case might tend to raise the average of the ratings for the second teacher. A third possibility is that the instructor may have two classes immediately succeeding each other in the same subject, so that he actually teaches better during the second period, owing to the fact that he has the material better organized, has had his attention called to points that were difficult for the first class, etc. So one might go on indefinitely listing possible factors. In general it is important to remember that judgments of traits represent not necessarily things possessed by the instructor, but rather the relationship existing between the raters and the one being rated. If one class happens to be made up of students very much interested in the subject matter quite apart from the teacher, while another class is a group for whom the subject is a necessary evil (compare students majoring in English with engineering students required to take English) the halo effect (see page --) will very likely tend to make the ratings of the interested group appreciably higher than those who lack interest.

How Nearly Do Members of the Same Class Agree?

Variations in the standard deviations of distributions of the various traits are also of interest. It is obvious, for example, that the relatively small spread of trait 1 for most instructors (see Tables III,

IV) is in sharp contrast with that of traits 5 and 10. In other words, students are more nearly agreed among themselves as to the relative amount of interest in his subject shown by the instructor than they are as to the quality of his subject-matter presentation, or the extent to which he stimulates intellectual curiosity in his students.

Other interesting comparisons are involved in the mean values of the traits relative to each other. We hazard the hypothesis on the basis of our data thus far that the average student rates the average instructor relatively high in "Interest in Subject-matter" and relatively low in "Presentation of Subject-matter." A plausible explanation of this tendency is that of the character mechanism of "projection." The student, in order to maintain his own self-esteem, rationalizes his own felt shortcomings by placing the blame upon the instructor's inadequate teaching. In general terms, the scale values for the different traits are not comparable.

It is obviously desirable to make possible the intercomparison of one's own standing as an instructor with that of an "average" instructor as a point of reference. While 100 percentism as an ideal is perhaps desirable on all traits, it is still true that in an imperfect world the individual instructor can perhaps more usefully measure his growth in grace in terms of a norm than in terms of a possibly desirable but practically unattainable absolute standard.

III—"HALO EFFECT"

In previous investigations of ratings by E. L. Thorndike, F. B. Knight, and others the so-called "halo effect" has come to light as a serious limitation. This effect is in brief the result of a lack of ability to discriminate among the traits of a rating scale for a given individual to be rated. The rater tends to judge the person to be rated in general terms as either relatively good or relatively poor on all desirable traits, owing probably to an emotional constant relative to the person in question. "The average rater," to quote Bingham and Freyd,[3] "will not rid himself of this bias and exercise his analytical powers unless the scale itself aids him to do so; consequently the absence of halo may be considered a point in favor of a rating scale."

We are forced to conclude, then, either that students in this situation are superior in rating ability to the "average rater," or else that our scale aids greatly in eliminating the halo effect. Table II gives intercorrelations on a number of traits for different instructors. It is important to remember that the lower these intercorrelations, the less is the halo effect and the more discriminative the scale. There probably should be some positive correlation in the sense that desirable traits are in general found to be positively correlated when measured objectively.

The intercorrelations are almost without exception so low as to be nearly without significance. The average r's of .258 and .248 for B and C respectively are not indicative of any considerable halo effect. Traits 1 and 6 (Interest in Subject-matter

[3]Procedures in Employment Psychology, A. W. Shaw Co., 1926, pp. 134-135.

and Sense of Proportion and Humor) even tend to be negatively correlated.

The point has been raised by some writers that such minute scale division as we, for example, have employed are useless, since the fine discriminations implied are in fact beyond the limits of human capacity. To this we may answer that since no one knows exactly *how* fine a discrimination is possible, it is best so to arrange the scale as to obtain the maximum discrimination of which raters are capable. Relatively minute scale divisions achieve this end.

Recapitulation

The experimental investigation of the Purdue Rating Scale for Instructors shows that student judgments have a considerable degree of reliability and since the Scale purports to measure only such student judgments, it is of considerable potential use for the self-improvement of instruction.

It is also established that the form of the Scale leads to no consistent errors in judgments so far as the location of the zero and 100 per cent ends of the Scale are concerned.

The halo effect—failure to discriminate among various traits of the person being rated—is a negligible factor.

There is considerable probability that the averages for the various traits for the same instructor are not directly comparable. Some traits are probably consistently rated lower than others for the average instructor.

Administering the Scale

In order to have conditions as nearly standard as possible, it is essential that the instructions at the

top of the Scale be carried out. On no account should the students be told the quantitative meaning of the Scale, for to do so would certainly produce results significantly different from what they would otherwise be. Nor, of course, should students, as previously stated, be asked to sign their names. This also would most certainly lead to distorted results.

Scoring the Results

The stencil attached to the Tabulation Sheet is designed to facilitate scoring. By placing this stencil directly underneath each trait scale, the numerical value of each can be read directly and entered in the right hand margin. By tabulating the results on the sheet provided for this purpose, the average results by traits can be very quickly obtained. The average of all the traits for any student is quickly obtained by adding the ten trait values and pointing off one decimal from the right.

Using the Results

The most valuable information will ordinarily come from an intercomparison of the average results for the ten trait values. If one finds, for example, that the average rating on the trait "Sympathetic Attitude toward Students" is low in comparison with those of other traits, a dispassionate self-analysis to discover just why this is so will very likely lead to a recognition of one's shortcomings in this regard. Having recognized the difficulty, one can set about discovering remedial measures. The section next following illustrates how one instructor used the results from his classes.

An Actual Case

The following tabular data were obtained by one young instructor for two different ratings by the same class. One rating was obtained a few weeks after the beginning of the semester, the second rating, near the end of the semester, after he had made conscious effort to change his reactions in class.

	Trait	First Rating Averages	Second Rating Averages	Average Improvement
1.	Interest in Subject	82.4	82.3	—0.1
2.	Sympathetic Attitude toward Students	64.4	64.3	—0.1
3.	Fairness in Grading	83.9	81.4	—2.5
4.	Liberal and Progressive Attitude	81.4	80.6	—0.8
5.	Presentation of Subject Matter	73.9	77.7	+3.8
6.	Sense of Proportion and Humor	65.5	80.9	+15.4
7.	Self-reliance and Confidence	65.9	80.5	+14.6
8.	Personal Peculiarities	68.8	81.4	+12.6
9.	Personal Appearance	87.4	89.8	+2.4
10.	Stimulating Intellectual Curiosity	73.0	79.4	+6.4

There is obvious and significant improvement in three traits. There is still considerable room for improvement all along the line, particularly in traits 2, 5, and 10.

TABLE I.—RELIABILITIES OF JUDGMENTS ON THREE INSTRUCTORS ON THE STRAIGHT FORM BY INDIVIDUAL TRAITS AND SUMS OF THE 10 TRAITS

Trait	Instructor T, N = 30		Instructor S, N = 33		Instructor R, N = 33	
	r	PE_r	r	PE_r	r	PE_r
1. Interest in subject	.70	.06	.57	.08	.61	.07
2. Sympathetic attitude	.65	.07	.92	.02	.43	.10
3. Fairness in grading	.53	.09	.46	.10	.61	.07
4. Liberal and progressive attitude	.68	.06	.51	.09	.56	.08
5. Presentation of subject	.47	.10	.55	.08	.69	.06
6. Sense of proportion and humor	.56	.09	.42	.10	.49	.09
7. Self-reliance and confidence	.69	.06	.51	.09	.44	.10
8. Personal peculiarities	.78	.05	.14	.12	.65	.07
9. Personal appearance	.90	.02	.42	.10	.80	.04
10. Stimulating intellectual curiosity	.60	.08	.55	.08	.63	.07
Sums of all traits	.83	.01	.50	.03	.64	.02

TABLE II.—INTERCORRELATIONS OF TRAITS. INSTRUCTORS B AND C

Traits	B (N = 52)		C (N = 52)		X (N = 31)	
	r	PE_r	r	PE_r	r	PE_r
1 vs. 2	+.140	.093	+.224	.092		
1 vs. 5	+.615	.062	+.298	.037	.19	.12
1 vs. 6	—.024	.095	—.177	.092		
5 vs. 10	+.141	.093	+.507	.061	.55	.09
6 vs. 9	+.196	.092	+.092	.094		
8 vs. 9	+.242	.091	+.325	.086		
2 vs. 3	+.498	.071	+.461	.076	.39	.10
3 vs. 4	-----	---	-----	---	.40	.10
Average	.258	---	.248			

TABLE III.—RATINGS BY 52 STUDENTS OF FOUR INSTRUCTORS ON THE TRAITS DESIGNATED AT THE LEFT OF THE TABLE

Instructor	A			B			C			D		
Trait	Mean	PE_m	σ_{dist}	Mean	PE_m	σ_{dist}	Mean	PE_m	σ_{dist}	Mean	PE_m	σ_{dist}
1. Interest in subject	70.15	2.10	22.30	96.83	0.33	3.54	89.04	1.19	12.70	86.25	1.32	14.15
2. Sympathetic attitude	79.90	1.31	13.90	91.54	0.84	8.50	72.89	1.82	19.50	56.83	2.08	22.20
3. Fairness in grading	86.92	1.33	14.10	94.50	1.11	11.25	87.41	1.63	17.40	85.19	1.58	16.90
4. Liberal and progressive attitude	66.33	2.68	28.40	92.40	0.79	8.45	73.75	1.52	16.25	81.16	1.47	15.75
5. Presentation of subject-matter	65.35	2.06	21.80	96.25	0.35	3.78	61.81	1.92	20.60	76.35	1.69	18.05
6. Sense of proportion and humor	75.54	2.14	22.70	93.30	0.77	8.25	53.56	2.16	23.15	76.93	1.65	17.70
7. Self-reliance and confidence	69.95	2.05	21.75	93.10	1.31	14.05	74.90	1.81	19.24	85.10	1.94	20.50
8. Personal peculiarities	82.21	1.78	18.90	89.20	1.54	16.10	81.70	1.81	19.40	69.62	1.94	20.50
9. Personal appearance	95.05	1.20	12.81	92.20	1.31	14.00	72.12	2.22	23.60	68.85	2.28	23.65
10. Stimulating intellectual curiosity	62.50	2.00	21.20	93.38	1.24	13.40	66.73	2.42	25.90	76.06	1.94	20.50

TABLE IV.—COMPARISON OF TWO FORMS ("STAGGERED" *vs.* "STRAIGHT") OF THE RATING SCALE FOR ONE INSTRUCTOR, N = 35

Trait	Straight				Staggered			
	Me'n	PE	σ_{distr}	Range	Me'n	PE_m	σ_{distr}	Range
1. Interest in subject	90.40	0.906	7.85	70–100	85.75	0.862	7.56	70–100
2. Sympathetic attitude	82.50	1.330	11.30	60–100	81.00	1.556	13.65	50–100
3. Fairness in grading	81.88	1.195	10.35	60–100	88.75	1.090	9.58	65–100
4. Liberal and progressive attitude	84.60	1.657	14.35	50–100	82.25	1.892	16.60	45–100
5. Presentation of subject-matter	60.20	2.230	19.95	0– 90	72.15	1.876	16.45	30–100
6. Sense of humor and proportion	76.25	1.940	16.80	30–100	67.25	1.340	11.76	45–100
7. Self-reliance and confidence	77.50	1.120	9.70	55–100	71.50	1.432	12.55	10–100
8. Personal peculiarities	76.66	1.980	17.15	40–100	81.75	1.553	13.62	60–100
9. Personal appearance	84.60	0.970	8.40	65–100	85.00	1.168	10.24	65–100
10. Stimulation intellectual curiosity	73.90	2.550	22.10	0–100	78.75	1.970	17.28	40–100
Total	76.04	0.366	9.75	45–100	77.75	0.445	9.82	40–100

TABLE V.—SIGNIFICANCE OF DIFFERENCES OF MEANS IN TABLE IV

Trait	Difference	PE_{diff}	$\frac{\text{Diff.}}{PE_{diff}}$	Chances in 100
1	4.65	1.25	3.72	99
2	1.50	2.04	0.74	69
3	6.88	1.59	4.33	100
4	2.35	2.52	0.93	74
5	11.95	2.91	4.12	100
6	9.00	2.35	3.84	99
7	6.00	1.82	3.30	99
8	5.09	2.56	1.99	91
9	0.40	1.51	0.27	57
10	4.75	3.22	1.47	84
Total	1.71	0.596	2.86	

Table V is to be read as follows: The difference of the means for trait 1 (Interest in subject) is 4.65 ± 1.25. The "index of significance" for this difference is given by the ratio of the difference to its PE, which is 3.72. The chances are 99 in 100 that this constitutes a "true" difference under the conditions of the collection of the data.

TABLE VI.—COMPARISON OF THREE CLASSES, INSTRUCTOR E

TRAIT	Class I, N = 34			Class II, N = 41			Class III, N = 25		
	M	PE_m	σ_{distr}	M	PE_m	σ_{distr}	M	PE_m	σ_{distr}
1	89.50	.724	.626	92.98	0.775	7.35	87.61	2.583[1]	18.00[1]
2	75.69	2.000	17.30	89.32	1.192	11.32	87.20	1.387	9.65
3	78.47	1.620	14.70	88.22	1.165	11.05	87.04	1.828	12.70
4	78.06	1.233	11.10	81.15	1.550	14.70	80.60	2.880	20.05
5	66.56	2.308	19.95	81.63	1.322	12.55	77.20	2.750	19.70
6	79.35	2.060	17.80	82.85	1.550	14.69	78.60	2.578	17.91
7	83.47	1.910	16.50	82.30	1.304	12.37	87.40	1.751	12.29
8	53.68	2.221	20.43	70.78	1.990	18.90	72.00	2.842	19.80
9	89.20	0.796	6.88	92.38	0.858	8.15	90.60	0.762	5.30
10	78.32	2.094	18.10	79.20	1.590	15.10	82.00	1.982	13.80
Total	77.230	0.362	3.125	84.081	0.228	2.180	83.061	0.432	3.205

[1]Relatively large values caused by one student's scoring 0 on this trait. But for this the values would be comparable to those of classes I and II.

TABLE VII.[1]—SIGNIFICANCE OF DIFFERENCES OF TOTALS IN TABLE VI

Classes	Difference	PE_{diff}	$\frac{\text{Diff.}}{PE_{diff}}$	Chances in 100
II-I	6.851	0.43	15.93	100
II-III	1.020	0.77	1.32	81
III-I	5.861	0.56	10.37	100

[1]To be read as in Table V.

Date Due

L. B. Cat. No. 1137

371.1 B81Y B20773
Bradenburg
Manual for the Purdue
371.1 B81Y B20773

DUKE UNIVERSITY LIBRARIES
Manual for the Purdue rating sc
371.1 B817
D90219388/